



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Hubris to humility

Citation for published version:

Isaac, A 2017, 'Hubris to humility: Tonal volume and the fundamentality of psychophysical quantities', *Studies In History and Philosophy of Science Part A*, vol. 65-66, pp. 99-111.
<https://doi.org/10.1016/j.shpsa.2017.06.003>

Digital Object Identifier (DOI):

[10.1016/j.shpsa.2017.06.003](https://doi.org/10.1016/j.shpsa.2017.06.003)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Studies In History and Philosophy of Science Part A

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Hubris to Humility:
*Tonal Volume and the Fundamentality of Psychophysical
Quantities*

Alistair M. C. Isaac

June 1, 2016

Abstract

Psychophysics measures the attributes of perceptual experience. The question of whether some of these attributes should be interpreted as more fundamental, or “real,” than others has been answered differently throughout its history. The operationism of Stevens and Boring answers “no,” reacting to the perceived vacuity of earlier debates about fundamentality. The subsequent rise of multidimensional scaling implicitly answers “yes” in its insistence that psychophysical data be represented in spaces of low dimensionality. I argue the return of fundamentality follows from a trend toward increasing epistemic humility. Operationism exhibited a kind of hubris in the constitutive role it assigned to the experimenter’s presuppositions that is abandoned by the algorithmic methods of MDS. This broad epistemic trend is illustrated by following the trajectory of research on a particular candidate attribute: tonal volume.

Not all facts are fertile.

Edwin G. Boring in 1962, reflecting on the
decline in interest in tonal volume.

1 Introduction

Can perceptual experience be measured? Committed to the view that it can, psychophysics attempts this measurement indirectly, by inferring a quantitative model of experience from the reports or induced behavior of subjects in response to physical stimuli. The question of whether this model constitutes a measurement, and whether one should infer from it that there are quantitative aspects of experience, has been vexed, to say the least.¹ Provisionally accepting these models as measurements, however, opens up a further array of interpretative questions. For instance: are all aspects of experience equally worthy

¹One source of criticism turns on the phenomenological point that sensations do not seem introspectively to be relatable by numerical ratios, but are rather purely qualitative (e.g. James, 1890, i.546–9). A second line of criticism takes psychophysics to mistakenly attribute properties of the stimulus to the experience, the “stimulus error” of Titchener (Boring, 1921). More recently, Michell, 1997, has argued that psychophysics has ignored the conceptually prior task of establishing that the qualitative axioms of quantitative structure are satisfied by sensory experience. See Marks and Algorn, 1998, 83–5, and Boring, 1942, 44–5, for surveys; Stevens, 1975, Chapter 2, gives a general response, while Barwich and Chang, 2015, provide a spirited rebuttal of Michell in particular.

of measurement? Or are some more “fundamental,” scientifically significant, or perhaps even *real* than others?

I argue that these questions were suppressed in early 20th century psychophysics, but reemerged later in the century as a natural consequence of new methods for extracting quantitative models from psychophysical data. Suppression was implicit in the operationist program for defining psychological concepts: if the scientific legitimacy of a sensory attribute² rests entirely on the existence of an operation that consistently assigns it values, there seems little room for assessing some attributes as more fundamental, significant, or “real,” than others. However, the rise of multidimensional scaling as a technique for extracting structure from data introduced a principled method for distinguishing fundamental attributes of experience from those that are merely derived or artifactual. Multidimensional scaling and related methods exhibit *epistemic humility* in the sense that they do not presuppose which (or even how many) attributes characterize a sensory experience. This contrasts sharply with the “hubris” of operationist psychophysics, which relied constitutively on presuppositions about the attribute of interest, as revealed most clearly in the instruction and training of subjects. This is ironic indeed given that operationism itself was responding to an earlier perceived hubris, and crisis, in psychophysics, turning in part on the vacuity of structuralist debates about the fundamental attributes. So, while fundamentality has oscillated in and out of theoretical importance, the trend toward humility has been largely monotonic: operationism embracing a more humble methodology than structuralism; multidimensional scaling emerging as a yet more humble alternative to operationism.

A full demonstration of this broad trend in sensory psychology is beyond the scope of this paper. Instead, I illustrate its influence in miniature through the example of a particular sensory attribute: *tonal volume*. Volume, the apparent “size” of a sound, emerges as an attribute of interest in the early 20th century, and in 1935 was offered by seminal operationist S. S. Stevens as a paradigmatic example of an “empirically meaningful” psychological concept. Nevertheless, interest in tonal volume waned throughout the 20th century, with more recent textbooks relegating it to passing comment, if mentioning it at all. By the lights of contemporary auditory psychology, volume is at best a quaint epiphenomenon, a historical footnote. I argue this rise and fall of volume is explained by the trend towards epistemic humility that allowed fundamentality to reemerge as a criterion of scientific legitimacy for sensory attributes.

The issue here is whether *within sensory experience* some attributes play a more fundamental role than others. Operationism, as practiced by Stevens and his advisor Edwin Boring, did allow that some attributes might be more basic than others, but not for reasons internal to sensory experience; rather, the relative significance of attributes could only be evaluated on physiological grounds. In the case of tonal volume, the fact that trained subjects cued to volume were able to consistently discriminate stimuli counted as evidence for the legitimacy of volume as a sensory attribute, but its relative importance compared to other attributes depended entirely on its conjectured physiological correlates. In contrast, multidimensional scaling derives fundamental attributes directly from psychophysical data. The fundamentality of these attributes is thus justified on purely sensory or behavioral grounds, independent of any physiological speculation. When applied to auditory experience, scaling methods fail to isolate volume as a fundamental attribute, hence its disappearance from the psychology of hearing.

²A note on terminology: “attribute” and “quality” were at one point used as contrasting terms, both of which Boring (1933) advocated replacing with “dimension,” appropriate if experience is treated geometrically. In order to emphasize the continuity in these issues from a modern perspective, however, I use the terms interchangeably, with a preference for “attribute” as it does not imply the contrast with quantity of “quality”—a contrast that would be inappropriate given that the psychological features at issue, even those labeled “qualities,” are treated quantitatively by psychophysics.

Section 2 illustrates the operationist suspicion toward fundamental attributes exemplified in the historical work of Boring, while Section 3 outlines the history of tonal volume. The puzzle of volume’s disappearance motivates an examination and critique of operationism in Section 4. I argue that the very logic Stevens employed to legitimate psychophysical measurement undermined operationism’s ability to differentiate between fundamental and derived sensory quantities, or to identify the inconsistent data on volume from different labs as an omen of its theoretical fragility. Section 5 introduces multidimensional scaling, explicating its criterion for fundamentality, and arguing it exhibits epistemic humility. When auditory experience is approached with this humility, volume does not emerge as a fundamental quality. I conclude with some new challenges for realism about psychophysical attributes that follow from this trend toward humility. The methodology of operationism succeeded in ensuring a firm logical link between sensory attributes and physiological mechanism; in contrast, multidimensional scaling’s enthusiasm for humility has left it without a clear foundational argument for how or why the fundamental sensory attributes it identifies are physiologically grounded.

2 Fundamental Attributes in Crisis

Edwin G. Boring’s (1942) *Sensation and Perception in the History of Experimental Psychology* is not only an important history of early psychophysics, it is also a document of prevailing attitudes toward that history during the rise of research on tonal volume. Boring presents Stevens’ research on tonal volume in the 1930’s as exemplifying a “new phase” in the way attributes of sensation are treated in psychology (25–7), the implications of which he had already explored and defended in his own research (Boring, 1933, 1935). What was the stagnant research program that this new phase moved to replace? Boring (1942) portrays a number of cases where debates about which phenomenal attributes were more “real” (23), “simple” (130), “primary” (375), or “basic” (512) turned out to be meaningless, poorly defined, or simply irrelevant.³ The most immediate target of this critique is the structuralist approach to psychology, but the breadth of Boring’s attack implies a more general position, that no evidence from experience, introspection, or behavior could establish theoretical primacy of one attribute over another.

For instance: one of the most notorious rivalries in the history of psychology was that between Helmholtz and Hering in the late 19th century. Although there were many facets to this rivalry, encompassing broad differences in theoretical commitments and methodology (Turner, 1994), one aspect that was perceived to turn on a point of fact concerned the number of primitive color sensations. Helmholtz argued for three—“red,” “green,” and “violet”—on the basis of his experiments on sensations of white induced by mixtures of spectral colors (Helmholtz, 1962 [1910], 143; c.f. Lenoir, 2006; Isaac, 2013). Hering argued for four—red, green, yellow, and blue—on the basis of phenomenological observations,⁴ for instance the color opponency effects revealed in experiments on afterimages (Hering, 1878, Report 6;

³Boring’s attitude was indeed representative of the time: for instance, the very same examples discussed here are presented as confused, irresolvable debates in the founding manifest of behaviorism, Watson, 1913. Stevens, 1935a, does not cite specific examples, but offers the primary motivation for operationism as the resolution of ill-defined, and thus “useless” disagreements, c.f. Hardcastle, 1995. Feest, 2005, 138–40, argues explicitly that Boring’s perspective on history shaped Stevens’ operationism.

⁴I’ve put the Helmholtz primitive colors in scare quotes, as for him the primitive qualities themselves may be “unconscious,” or at least difficult to directly access in experience without training—the names for these primitive sensations are thus placeholders, derived from the experiences associated with the single wavelengths of light that most excite each to the exclusion of others, although such exclusive excitation is impossible in practice as the receptive fields for the corresponding receptors overlap. In contrast, Hering’s primitive colors, in virtue of being directly accessed in phenomenological observation, unambiguously refer to everyday notions of red, green, etc.

Turner, 1994, 130–4, 189–95); this position was then bolstered by work showing subjects are able to judge the degree of purity of the four Hering primaries.⁵

For Boring, there are two aspects to this controversy. On the one hand, there is a legitimate question about the neural correlates of behavior that receives a straightforward physiological resolution. Once we recognize that not all physiological processes relevant to a sensation need reside in the sensory organ—three processes corresponding to Helmholtz’s primaries may reside in the retina, and, without contradiction, four (rather, two opponent) processes corresponding to Hering’s primaries may reside in the brain—the dispute disappears. On the other hand, there is the question of which qualities of experience are simple or basic to its phenomenal character. Hering and Helmholtz themselves conceived of this as the question at issue: which color sensations are primitive, determining through their combination all other color sensations? Out of their initial dispute evolved an elaborate debate about the number of simple color sensations, and whether particular sensations are simple or complex. Boring is completely dismissive of this aspect of the issue, portraying it as ill-defined and confused:

That the concept of a simple color had become ambiguous is obvious. It is one thing to say that there are seven different colors or four pure colors or three elementary colors, and quite another to say that there are 150 discriminably different colors. . . . The problem was never settled. It simply disappeared. From the perspective of the present it is clear that the proposition was never clearly formulated. Since simplicity and complexity were supposed to be immediately apparent to introspection, no rigid criteria for either concept were ever put forth so that the issue could be determined, or at least be given sense with reference to definitions of *analysis*, *complexity* and *simplicity*. (130–1)

More generally, Boring countenances sensory attributes only so long as clear experimental criteria exist for distinguishing them, and if such a criterion exists for an attribute, he infers it may plausibly be identified with a physiological process. The debate on attributes general to sensation as a whole, for instance, he portrays as a struggle from introspective speculation toward experimental criteria (19–27). The introspective endeavor attempted to enumerate basic components, or “elements,” in experience, and struggled with the question of what to treat as elementary, and what to treat as attributes of these elements (as in the structuralism of Titchener, 1896, and Wundt, 1907). Thus, for instance, vision has the “quality” of color, plausibly treated as elementary, but it is questionable whether it possesses the classic “attribute” of intensity (21). Once it is clear that only “attributes” may be probed through experiment, however, these questions fade away. Thus, Boring appears to endorse the conclusions of Rahn (1913), which he summarizes:

Every psychophysical experiment on sensation is accomplished under some set to observe and to judge a particular attribute. All that the observation shows is the attribute, which must, therefore, be regarded as the *observed datum*. The sensation, on the other hand, consists of all the attributes which might have been observed under all the possible instructions; in other words, it is a *physiological entity*, a total excitation which carries with it these potentialities for attributive report. (24, emphasis in the original)⁶

⁵Surveys on pure (or “unique”) hues: Kuehni, 2014; Wright, 2016.

⁶I believe this passage is more reliable as a representation of Boring’s views than of Rahn’s—the basic ideas are there in Rahn, 1913, but the emphasis on the physiological is not the same; c.f. Section 4.

The elements of Titchener were fundamental in the sense of being basic and indivisible building blocks of thought (or “mental experience,” 1896, 13). In rejecting a distinction between “elements” and “attributes,” Boring likewise rejects a criterion for fundamentality of perceptual quality grounded solely in psychophysical measurement. Insofar as there remains a criterion for more or less fundamental attributes, this criterion rests entirely on the physiological basis for those attributes (Boring, 1933, 122). This project is not discontinuous with that of Titchener, but it transposes his concern with mental structure into the domain of physiological processes (ibid., vii; Feest, 2005, 139–40). This transposition preserves sensory attributes as legitimate objects of scientific study within a post-behaviorist psychophysics, for which “The immediate task . . . [is] the understanding of the physiology . . . of consciousness” (Boring, 1933, 31).

A case in point is the attribute of *extensity*, spatial spread or extent. Titchener (1924) lists extensity as a fundamental dimension of experience because it is revealed through introspection and behavioral studies across multiple sensory modalities. Boring (1933) also accepts extensity as an attribute of particular interest, not for its phenomenal character, but because its corresponding physiological process may be easily conjectured, perhaps the projection of extended stimulation at the periphery to a spatially extended region of the brain (67–74). In particular, the auditory manifestation of extensity, *tonal volume*, is an attribute of interest because it provides evidence for a physiological process analogous to that associated with extent in vision, yet in the organ of audition (Boring, 1933, 85; 1935; Stevens, 1934).

3 A Brief History of Tonal Volume

The story of tonal volume is simple: volume emerges as a distinct psychophysical quantity; a litany of standard psychophysical properties of volume are measured; volume is forgotten.⁷

The psychophysics of sound, as with all psychophysics, begins by studying our perceptual experience of the simplest possible stimuli. The (physically) simplest sounds are “pure tones,” i.e. sine waves. A sine wave stimulus that begins and ends abruptly may be defined by just three parameters: *amplitude*, *frequency*, and *duration*. The degree to which the stimulus disturbs the surrounding medium, and thus the strength with which it impacts the subject, i.e. its *energy*, is a function of both amplitude and frequency.⁸ The most basic experiments typically control for duration, so only energy and frequency were deemed relevant for reporting many of the results discussed here—a point we revisit later. The two sensory attributes that appear to track energy and frequency in the stimulus are *loudness* and *pitch* respectively. *Volume* is a third attribute, apparently present even in the perception of pure tones.

In contrast to our colloquial use of the terms “volume” and “loudness” interchangeably for apparent sound energy, “tonal volume” is used in psychology to characterize the apparent spatial expansiveness

⁷Other surveys of the psychophysics of volume identifying many of the same milestones: Boring, 1942, 378–9; Licklider, 1951, 1004–5, 1019; Gulick, Gescheider, and Frisina, 1989, 283–8; Hollander, 1994, 21–4.

⁸In order to clarify discussion, I have decided to reserve “energy” for the strength of the stimulus and “intensity” for the strength of the perceptual response; in the case of sound perception, “intensity” is called “loudness.” This decision is somewhat arbitrary and glosses over several terminological complexities. One problem is that both the acoustical and the psychoacoustical literatures sometimes use “intensity” to refer to a property of the stimulus, but this introduces a dangerous ambiguity as stimulus intensity and experience intensity are not equivalent. Since a clear distinction between properties of the stimulus and properties of experience is absolutely critical for understanding the methodologies of *both* operationism and multidimensional scaling, I have opted to suppress as much as possible this practice. A second issue is that acoustics employs several different measures of the energy of a disturbance in a medium: total energy of the system; the power, or rate of energy consumption; the flux, or rate of energy flow across a surface; the pressure exerted on a surface averaged over time; etc. Not all studies on sound perception employ the same measure of stimulus energy—since these are typically intertranslatable, however, I have also suppressed this detail. For a clear survey of the acoustics necessary for auditory psychophysics, see Gulick, Gescheider, and Frisina, 1989, Chapter 2.

of a sound. James (1890) observed that loud sounds also seem to have a great spatial extent (ii.136). Titchener (1908), quoting James, posited that volume should be taken as a “qualitative attribute” of sound perception, on par with pitch (12–16)—and in contrast to the intensity of a sound, i.e. loudness.

These, however, were mere conjectures, and the question of whether volume should truly be treated as an independent attribute of sensation, distinct from pitch and loudness, was left open to future experimentation. For Titchener (1908), this question turned on whether volume could be varied independently of the other attributes. James had already noted that volume seems correlated with loudness, while Stumpf (1883) had observed that it correlates with pitch—low tones sounding large and high tones sounding small.⁹ Thus, the question of whether volume constituted a legitimate attribute of sound perception appeared to turn on whether an experiment could be devised in which volume varied, while pitch and loudness remained constant.

A first step was made by Rich (1916), who demonstrated that the difference limens for volume differ from those of pitch. This means that tones sensibly different with respect to pitch may nevertheless be indistinguishable with respect to volume; consequently, even if volume correlates with pitch, its rate of change with respect to change in stimulus frequency is different. This was a quantitative result, but Rich also collected qualitative reports from his subjects, with the intent of establishing that volume perception was not merely a matter of metaphor or association (we return to these reports in Section 4). Next, Halverson (1924) established a similar result for the limens of volume and loudness.¹⁰ By varying the amplitude of the stimulus while holding pitch constant, he was able to establish that the rate of change in volume differed from that of loudness, even though both varied in the same direction, namely increasing with amplitude.

The definitive experiment for disentangling volume from loudness and pitch was finally executed by Stevens (1934). As emphasized by Boring (1942), Stevens obtained this result by changing the criterion for distinguishing attributes from that of Titchener: rather than independent variability, independent constancy was taken as the hallmark of a legitimate attribute (25–6). Although it is not possible to hold both loudness and pitch constant while volume varies, it is possible to hold volume constant while loudness and pitch vary. Furthermore, in obtaining this result, Stevens discovered a fourth attribute of the sensation of simple tones, an apparent “compactness” or “concentration” of the sound, which he called *density* (458). The fundamental reason that constancy is the relevant criterion is that all four attributes are distinct functions of *both* the properties that define the stimulus (energy and frequency); thus variations in either of these physical quantities will simultaneously affect the values of all four sensory attributes. Figure 1 shows curves of constancy for each of Stevens’ four attributes graphed against the two properties of the stimulus.

After Stevens’ demonstration that volume was indeed an independent attribute of tonal perception, further features of volume, of the kind measured for other perceptual attributes (loudness, pitch, color) were systematically investigated. Thomas (1949) extended the determination of contours of equal volume (against values of energy and frequency) to a range far beyond that mapped by Stevens. He trained naïve

⁹Of course, Stumpf didn’t use the term “volume,” but he remarked that lower tones appeared to have a larger expanse in perception: “Den tiefen Tönen kommt in der Vorstellung eine grössere Ausdehnung zu” (207); c.f Boring, 1942, 378.

¹⁰An earlier study (Halverson, 1922) had shown that volume varies with difference in phase between binaurally presented stimuli. While Halverson was able to establish there was no sensible difference in pitch, thus confirming Rich’s result that volume limens are not equivalent to those of pitch, the study was unable to establish the exact relation between apparent differences in loudness and in volume while phase varied. Odenthal, 1963, also demonstrated the change in size of an auditory “image” with changes in phase, although he is unconcerned with distinguishing volume from loudness.

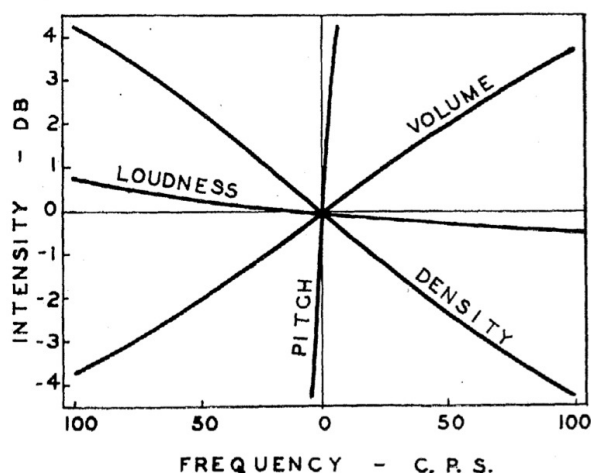


Figure 1: Constancy curves for the four attributes of the sensation of simple tones (Stevens, 1934, 458); “intensity” refers to the property of the stimulus called “energy” in the text (see Footnote 8).

subjects to match tones in volume by both adjustment of frequency and adjustment of energy. Thomas offers two considerations against the worry that volume assessments do not concern a truly independent attribute, but are merely “a derived or mediate judgment on the basis of relative positions upon the two absolute scales of pitch and loudness” (197). First, subject’s introspective reports do not describe any such mediate procedure; second, plotting equal volume contours against the scales for pitch and loudness (properties of experience, rather than the frequency and energy that characterize the stimulus) does not reveal any simple functional relationship (197–8).

These results characterize the topological structure of volume, providing a topographic map of volumetric sameness and difference, but without any quantitative assessment of distances within this landscape. A first attempt at determining the metrical structure of volume was made by Terrace and Stevens (1962), who used the method of magnitude estimation to determine distances in volume from an arbitrary standard to various comparison tones. Subjects were instructed to attend to volume (rather than pitch or loudness) and then played a reference tone, which they were told had a volume of “10”; subjects then assigned numbers to subsequent stimuli, chosen from Thomas’ contours. The relative agreement across subjects and trials in the numbers assigned to stimuli falling on each of Thomas’ equal-volume contours confirmed his results, and the difference between averaged numerical values for different contours motivated a tentative assignment of distances between their respective volumes. These results were then supplemented by Stevens, Guirao, and Slawson (1965), who used the same method to assess the quantitative relationships between loudness, volume, and density, concluding that loudness is proportional to the product of volume and density. Gulick (1971) employed a different method for quantifying subjective volume; his subjects adjusted the frequency of a comparison tone against a standard to produce half and double the relative volume. By varying amplitude to ensure energy remained constant, Gulick was able to produce a “subjective scale” of volume distance as a function of frequency (160).¹¹

¹¹See Isaac, 2013, for a discussion of equivalent experimental methods employed in the transition from merely topological to metrical representations of color experience.

The later volumetric studies of Stevens and Gulick are exceptions, however. After the flurry of activity leading up to Stevens (1934), interest in volume waned. For instance, Carterette and Friedman (1978) devote a single sentence to volume, while contemporary introductions such as Goldstein (2007) or Moore (2008) don't mention it at all. Why the decline in interest? How could a sensory attribute deemed exemplary by leading psychophysicists fall so rapidly and thoroughly from grace?

Attempting a revival in tonal volume research in the early 1980's, Perrott and colleagues offer two possible explanations for this decline. Anecdotally, they claim the most popular hypothesis about tonal volume is that it simply reflects a learned correlation between sound quality and the size of sound sources, but "Volume as a *learned* attribute would be of little interest to researchers studying primary features of the acoustic spatial process" (Perrott and Buell, 1982, 1413). Alternatively, Perrott, Musicant, and Schwethelm (1980) trace the decline in volume research to its very heyday in the 1930's. Even as Stevens was publishing his landmark results, other labs claimed that volume research could not be replicated. Gundlach (1929) failed to recover Halverson's results, with subjects' reporting the "impossibility" of matching distant pitches in volume (194); Gundlach and Bentley (1930) also report that the limens measured for volume do not agree with those of Rich and Halverson (530). Zoll (1934) supplements the measurements of Rich, Halverson, and Gundlach with his own, demonstrating "instability" in the values of volumic limens, both across labs and across subjects in different conditions. He concludes that, since "[t]he measure of the value of a concept is its stability under varying experimental conditions" (99), volume is not a legitimate scientific concept.¹²

Gulick (1971) dismisses these results as useless, since "insufficient account was taken by these experimenters to the instruction of listeners" (158)—a response that reflects the operationist methodology advocated by Stevens (see Section 4). Perrott, Musicant, and Schwethelm (1980) offer a different diagnosis, attributing the divergence in results to the experimenters' failure to control for stimulus *duration*. Intrigued when a subject on another experiment reported that the sound of a stimulus "gradually grew until it appeared to fill his whole head" (44), they investigated the effect of duration on volume perception. Furthermore, in an attempt to disambiguate the dimension-relative notion of "extent," and the prejudice toward tridimensionality implicit in the term "volume," they asked subjects to attend to the *width* of the stimulus. They were able to demonstrate tonal width assessments by untrained subjects increased with duration of the stimulus, concluding "the apparent size of an auditory image is dynamic" (55). While this result sparked new research on sound and size, the ensuing studies have taken an ecological focus, probing subject success at inferring spatial characteristics of a sound source in complex environments (e.g. Rudno-Rudziński and Renowski, 1990; Hollander, 1994; Potard and Burnett, 2003; Grassi, 2005). Fascinating as this research may be, it does not address the question of whether volume is an intrinsic attribute of auditory experience.

While there is some truth to the considerations offered by Gulick and Perrott, I will explore another possible explanation for the decline of interest in tonal volume here, one suggested by Carterette and Friedman (1978) immediately after their very brief reference to volume:

We [perceptual psychologists] now examine "attributes" as parameters or dimensions using pattern recognition and multidimensional scaling methods. (15)

¹²For a broader survey of disagreement on volume in the 1930's and early 40's, see Coffin, 1941, 83–5, who takes this as evidence of the "ambiguity" of the concept. Coffin himself contributes to doubts about the legitimacy of volume by showing it is highly susceptible to suggestion (96–9).

I argue in the following two sections that with this change in methodology came a change in attitude toward the importance and determination of fundamental perceptual attributes. Volume is a casualty not of idiosyncrasies in the design and reporting of particular experiments (as Gulick and Perrott would have it), but of a discipline-wide shift in values. From this perspective, the lack of contemporary interest in volume was foreshadowed in the very features that ensured its operationist legitimacy: the ineliminable role of the experimenter in directing subject attention toward volume and the simple mathematical relationship obtaining between volume, density, and loudness.

4 Operationist Hubris

The aim of this section is to articulate the role of Stevens' operationism in the rise, and subsequent fall, of tonal volume. After outlining the basic logical structure of operationism, I demonstrate the constitutive role of the researcher's presuppositions in operationist psychophysics—its “hubris.” Stevens argues that the legitimacy of a psychological concept depends on ensuring through ostension and training that subjects respond to its stimuli as the experimenter intends—an attitude that emerges as the methodological constant across all successful studies on volume. The critical role of ostension and training in these experiments is illustrated through a closer look at the reports of subjects, who express a diversity of subjective experiences, only gradually shaped through training into a consistent pattern of behavioral response. I next rehearse Boring's argument that a direct consequence of Stevens' methodology is the potential for an arbitrarily large number of perceptual attributes of equal theoretical interest. This will return us to Stevens' result of 1965—a result that reinforces volume's operationist legitimacy, but will damn its chances at fundamentality by the logic of multidimensional scaling.

Operationism arose in psychology in the late 1930's and early 1940's as psychologists realized the strict behaviorism of Watson was inadequate for capturing the full range of psychological phenomena, yet before Skinner emerged as a crucial figure and the second wave of behaviorism gained momentum. Thus in 1942 Boring was able to write of behaviorism in the past tense, stating it “ultimately disappeared” (12), and portray operationism as its successor. Even before the christening of operationism, self-professed behaviorists articulated its basic tenet, that mental states are legitimate objects of scientific study insofar as they may be detected by objective, empirical procedures. As early as 1926, for instance, Edward Chance Tolman had argued that even “conscious-awareness” could legitimately be detected in objectively described and publicly available patterns of behavior, an attitude he would later go on to describe as operationalist (Tolman, 1936; c.f. Amundson, 1986; Feest, 2005).

Stevens articulated the fundamental tenets of his operationism in two articles of 1935. Stevens (1935a) outlined the basic idea, that psychological concepts may be given scientific validity insofar as they are defined in terms of the operations, or procedures, performed to detect them. By defining concepts in terms of operations, psychology may avoid “hazy, ambiguous and contradictory concepts,” and thus “silenc[e] useless controversy” (323). Stevens (1935b) reiterates the basic claims of the previous article, emphasizing the importance of the “discriminatory response” as the ultimate arbiter of conceptual legitimacy (522). He then illustrates operationist methods in the study of tonal attributes, in particular the discovery during his research on volume of the attribute of density.

Stevens' emphasis on discriminatory behavior reveals an implicit, but important, inferential step in his operationist reasoning. Psychophysicists attribute the ability to discriminate only if the subject reliably

and consistently makes a distinction across repeated trials—because perception is inherently statistical in character, repeated probing is required to factor out the effects of noise and randomness in behavioral responses to perceptual stimuli (an insight canonized by Fechner, 1860). So, an operationist may “define” a psychological concept in terms of an operation to detect it, but it is only when this operation has proven its power to produce consistent data, i.e. a genuine discriminatory response, that the psychologist may infer its scientific legitimacy. The referent of that concept then “exists” in the sense that it is present as an “object or event,” i.e. that defined by the procedure, and no linguistic debate about the “adequacy” of the “definition” may impugn that existence (1935b, 518–9). For instance, volume may be inferred to “exist” as a legitimate psychological phenomenon because experimenters may train subjects to attend to volume, and once they are trained, subjects consistently discriminate the sameness or difference of tonal stimuli with respect to volume.

The history of density, Stevens’ (1935b) example of an operationally legitimate psychological concept, is the history of volume writ small. It begins with the experimenter’s observation that some subjects were equating tonal stimuli “in a consistent and unique manner, characteristic of none of the other known attributes” (526)—just as Stumpf and James had observed patterns in our intuitive, pre-theoretic assessment of tonal “size.” This counts as evidence that a legitimate concept is nearby, but is not itself enough to determine the procedure for defining that concept:

[N]ext the problem arose as to what operations would be required to show that the concept of tonal density could be extended to a naïve observer. First of all we could merely ask him to notice the density of a tone. If his past experience had been of the right sort, it might happen that this instruction would be sufficient to ‘tune’ him to the proper attitude. (In practice this usually turns out to be the case.) Suppose, however, he had had no experience with the word *density* in any of its usages, then we might have to dispense with language and resort to a round-about procedure which may be called the method of *successive approximation*. . . . We should keep the subjects at work by various motivating devices and reward them only when they achieve the reaction for which we are looking. (526)

Stevens’ assertion that past familiarity with the term “density” is “usually” enough to ensure the “proper attitude” in the subject misleadingly implies that the subject’s antecedent understanding of tonal density is the object of study. Strictly speaking, however, verbal instruction of this sort is only methodologically sound if it produces behavior in accordance with the *experimenter’s* understanding of the attribute at issue. As Stevens (1935a) makes clear, rigorous operationism may never rely solely on the subject’s understanding of words like “density,” for

If the experimenter assumes that the words used by the subject mean what they would mean if the experimenter used them, he is apt to be in error, for the meaning of a word depends upon the past history of the person acquiring the meaning, and no two personal histories are identical. (329)

Strictly speaking, without knowing the exact personal history of the subject, there is no way to know that they use words in the same way as the experimenter him or herself (330). Thus, it is ultimately only through the sameness of evinced behavior that the sameness of concepts can be assured, and the experimenter must “tune” the subject to produce the desired behavior in the same manner as animal subjects are trained: through ostension of exemplars and reinforcement.

[I]f it is a question of conveying what is meant by a musical fifth to a completely naïve person, the most satisfactory procedure is to sound two tones in the proper way and to say, “*that* is what is meant by a fifth.” (324)

In principle, if this step of is omitted, and subjects are instructed to attend to some theoretical attribute—“volume,” “density,” “a musical fifth”—without ensuring their (behaviorally evinced) understanding of the term matches that of the experimenter, and thus also of the other subjects, data gathered from them could not be meaningfully aggregated or compared. This data would be infected by the possibility that each subject attended to a different attribute, picked out by the semantics of his or her idiolect.

The role of ostension by the experimenter, in order to ensure the subject’s responses match the one the experimenter is “looking for,” is thus essential to the operationist project. It follows as a consequence of operationism’s conflicting tenets that (i) the meaning of concepts for any subject is a consequence of his or her personal history, and thus *de facto* private and (experimentally) inaccessible, yet (ii) “Science demands public rather than private facts,” which must be “independent of the observer” and “which all normal men can observe alike” (327). The solution to this apparent conundrum is the hubris of operationism: concepts must originate with the experimenter, and subjects must be “tuned” to the experimenter’s prior understanding through training, on pain of incommensurability.

The need for a specialized “tuning” of the subject is amply illustrated in the history of tonal volume. Operationists such as Boring and Stevens insisted that judgments of volume were naturally and easily made, yet at the same time acknowledged the need for subject training:

[I]t is a very simple matter to convey to a naïve observer what is meant by volume and density by presenting him with a high tone at about 4000 cycles and asking him to note its ‘smallness’ (volume) and its ‘compactness’ or ‘concentration’ (density) and to contrast these characteristics with those of a low tone at about 200 cycles. The difference is at once obvious. (Stevens, 1934, 457–8)

Yet the discussions of those beginning from a position of greater skepticism about the legitimacy of volume as a tonal attribute are quite different. Rich (1916) ultimately concludes that volume is a legitimate attribute from the agreement in differential limens he is able to obtain; however, the introspective reports he collects from his subjects betray a far more complex phenomenon than that indicated by Stevens. Crucially, the subjective procedures for assessing volume differ across subjects. One employs a kind of “imitative kinæsthesia,” imagining the shape of his mouth or chest while singing the tones (18); while a second occasionally employs kinæsthetic criteria, but more commonly relies on visual imagery, directly reporting an inverse correlation between volume and density (50 years before Stevens, Guirao, and Slawson, 1965): “The larger tones seem to have a greater diffuseness. It seems as if the smaller ones fill space more compactly, what they fill.” Only later in his training do the visual criteria “drop out, and judgment becomes more direct” (19). Rich’s third observer appeals frequently to the size or “capacity” of sound sources—bigger tones “come from a bigger vessel”; while he initially also describes a sense of being filled by the tone, this drops out: “Now I use as criterion the size of the vessel emitting the sound, and I seem to imagine the sound visually, as spread out behind me” (20).

These reports portray a diversity of phenomenal experiences while assessing volume that, while they ultimately converge to a (fairly) uniform pattern of response, do so only through directed exercise. Does the subordination of individual differences during this training regimen undermine the operationist

inference from consistent discriminatory response to scientific legitimacy? Not if one takes the *potential* to behave consistently as itself a psychologically important fact. This is Boring's (1935) *apologia* of Stevens, that the subject's ability to discriminate is evidence of a neural "differentiation," and such differences of neural response are evidence for an adequate theory of perception (241–2). This is essentially the same position he attributes in 1942 to Rahn (1913): differential responses under different instructions are simply means of accessing "a *physiological entity*, a total excitation which carries with it these potentialities for attributive report" (24).

Boring (1935) argues that Stevens' results show it is incorrect to conceive of the attributes of sensation as standing in one-to-one correspondence with properties of the stimulus. Pitch is a function of *both* frequency and energy, not merely frequency; likewise, loudness, volume, and density are each, also, distinct functions of both frequency and energy. This result refutes the sensory realist's presumed isomorphism between the physical world and the world of experience; consequently, there is no longer any reason to suppose that a stimulus defined by n parameters will result in an experience characterized by n (or less than n) attributes. But this result has a corollary: there is no principled limit on the number of attributes ascribed to a sensory experience:

Theoretically there is no upper limit to differentiation. A stimulus of one, two, three or m dimensions can have one, two, three or n attributes in the resultant sensation.

The only limits are "practical." On the one hand, practical limitations in the organism:

Differentiation of the sort that can lead to differential neural reaction is not unlimited in the finite organism. One expects the basis for a few attributes but not for many.

On the other hand, practical limitations on the subject and the experimental method of discrimination:

[S]uppose that the neurology of hearing were such as to form the basis for twenty different isophonic contours in the audiogram of [Figure 1]. Should we then expect to be able to plot all these contours? Hardly, for discrimination would break down when differentiation between functions becomes so slight... It is always easier to discriminate the few than the many. (245)

So, by the lights of operationism, as many attributes may be ascribed to a sensation as may be consistently measured, i.e. assigned curves of constancy on the basis of discriminations. There is no principled upper limit to the number of these attributes. Furthermore, neither properties of the stimulus, nor aspects of behavior provide a principled reason to value one over the other. In the former case, for instance, one might argue that loudness is more scientifically significant than volume as it corresponds to a significant physical feature of the stimulus, the energy, while volume does not; however, once we accept that both volume and loudness are functions of both energy and frequency, this argument is blocked. In the latter case, one might argue that loudness is more scientifically significant than volume since it is more salient, or discriminated more easily by naïve subjects; once one grounds the legitimacy of perceptual attributes in the mere potential for evoking a consistent discriminatory response, however, this argument is also blocked, as both sorts of response may be demonstrated.

The upshot of this line of reasoning is conceptual egalitarianism: no attribute is more fundamental than any other. In the case of Stevens' four attributes of tone, this ontological symmetry is mirrored in their mutual interdefinability, since "any two ... determine the other two when the laws of relationship are

known” (240). It is in regards to this apparent symmetry that Stevens, Guirao, and Slawson (1965) is so important. They demonstrate that loudness is proportional to the product of volume and density, “as the analogy with physical density . . . suggest[s]” (507). On the one hand, this result makes significant progress toward the “law” that links the four tonal attributes, and thereby provides a further confirmation of the legitimacy of volume by operationist lights; on the other, however, it reveals a kind of logical asymmetry: loudness, volume, and density are closely and simply related, while no simple formula for deriving pitch from any two of them is known.

Intuitively, there is a kind of redundancy, or ontological profligacy, here. Of the four attributes, loudness, pitch, volume, and density, any two may be derived from any other two. Furthermore, amongst loudness, volume, and density, the relationship is simple and known. Why not simply focus attention on pitch and one of the remaining three attributes? More generally, why countenance many attributes if few are enough to specify the sensation? The theoretical basis for this line of critique, and the methodological alternative it suggests, are the topic of the next section. However, it is worth briefly considering the fate of the operationist answer.

The operationist’s resolute physicalism maintained that the potential to consistently discriminate a tonal attribute was evidence for some neural differentiation during auditory processing. Stevens (1934) hypothesizes that volume may be correlated with the spatial breadth of excited fibres on the organ of Corti.¹³ Boring (1933) gives a similar explanation, positing the extended attribute of tones is derived from the spatial extent of excitation on the organ of Corti, and points out that our failure to notice it outside the laboratory may be because it has “little practical use” (85). This physiological explanation was revived by von Békésy (1963), who argued from analogies between his theory of the physiology of the cochlea and experiments he had performed on touch sensations of vibrating bodies. However, while von Békésy’s theory of cochlear mechanics has turned out to be essentially correct, his hypothesis concerning the correlates of volume perception appears to have led nowhere. In particular, in the physiological literature (as in the psychoacoustical), as attention has turned to perception of the spatial location and size of sound sources, literal extent as a feature of the auditory signal has dropped from significance. Thus, the supposed physiological fruits of operationist volume research were never borne.

It seems then that the moral of Rahn (1913), in whom Boring saw a precursor to operationism, should have been read as a warning:

When his interpretation of the “world” fails the philosopher, he may come back to the “facts of sense” from which to construct the world anew, but he gets from sense only that which comports with his purpose and his problem. (126)

5 Multidimensional Scaling: Humility and Fundamentality

Multidimensional scaling (MDS) is a method for producing a low-dimensional representation of high-dimensional data.¹⁴ This section examines a particular instance of the method, proposed by Roger

¹³Transduction of the mechanical energy of the sound wave into neural energy takes place in the cochlea, a fluid-filled tube in the inner ear, wound into a snailshell shape. The organ of Corti is on a membrane stretched along the length of this tube, and bears cilia (fine hairs), that stimulate the dendrites of auditory nerves when a wave disturbs the fluid.

¹⁴Although I focus only on Shepard’s version of MDS, it is representative of a wide array of techniques (including, e.g., Factor Analysis and Principal Component Analysis) that, while mathematically quite different, exhibit the same epistemic features. The motivation for focusing on Shepard’s 1962a and 1962b in particular is their historical significance, and the clear and reflective discussion they contain.

Shepard in two influential papers of 1962. The popularity of MDS (and related methods) in psychology has grown steadily since this proposal; although typically omitted from introductory texts, perhaps due to its mathematical sophistication, its importance for actual research is attested by the wide array of survey articles and handbooks devoted to it.¹⁵ I cannot pretend to do justice to the mathematical details of MDS here, nor to survey the many methods that share its basic epistemic perspective. Rather, the goals of the present section are to introduce enough of MDS to demonstrate its epistemic humility, to articulate its principled criterion for the fundamentality of perceptual attributes, and to explain why tonal volume fails this criterion.

Contemporary psychophysicists who employ MDS are engaged in the same broad theoretical project as the operationists, the structuralists, and the 19th century pioneers of basic psychophysical methods: the identification of attributes of sensory experience, and the representation of these attributes in a quantitative model. These models often take the form of geometrical spaces, the most familiar perhaps being the color solid. Although typically referred to as “quality spaces,”¹⁶ since they organize sensory qualities, these models are *quantitative* in the sense that they may also include a metric, or measure of distance between particular “sensations,” i.e. points in the space—for instance, the measurements of distance between tonal volumes made by Stevens and Gulick discussed in Section 3.

Since MDS is a method for analyzing data, it does not directly compete with the experimental methods proposed by any earlier school; data from operationist or even structuralist experiments could in principle be analyzed through MDS. Nevertheless, MDS encourages a very different perspective on the collection of data. Whereas operationist methodology demands that the experimenter impose a criterion for assessing stimuli on the subject through ostension and instruction, an experiment intended for MDS need only instruct subjects to compare stimuli *somehow*. The criteria used by the subject to compare stimuli is still the topic of interest, but using MDS it may be extracted from the data. Just as with operationism, there is an inferential step here: only if subjects discriminate stimuli consistently may a legitimate sensory attribute be inferred. However, the operationist emphasis on the *possibility* of post-training discrimination is replaced by MDS with a focus on how subjects antecedently *do* discriminate.

MDS may be applied to any proximity data, i.e. collection of items with some assignment of “closeness” between every pair. In the case of perception, the relevant notion of “closeness” is typically understood as similarity: orange is more similar to red than to green; some people find peanut butter tastes similar to sesame paste, others do not; the weight of these two objects is so similar, I can’t tell which is heavier. We can elicit these judgments in a psychophysical experiment by presenting stimuli from the same domain (color patches, foodstuffs, boxes of different weight, etc.) and asking subjects to order them, distinguish them, report verbally on their properties, or otherwise demonstrate through behavior the degree of perceived similarity between them. These are the kinds of judgments that historically have been used to determine perceptual limens and the topological relations within a quality space; they are also instances of the kind of discrimination that Stevens took to ground operationism. Viewed very generally, these judgments generate proximity data in the form of a qualitative ordering over pairs of stimuli by their degree of similarity.

Proximity data may also be numerical, and thus *prima facie* quantitative, for instance if an exper-

¹⁵e.g. Carroll and Arabie, 1998, who themselves cite a litany of other surveys; Takane, Jung, and Oshima-Takane, 2009; Ding, 2013; Borg, Groenen, and Mair, 2013.

¹⁶For a philosophical introduction to quality spaces, see Clark, 1993; Boring, 1933, calls them “qualitative continua” and foreshadows the general geometrical perspective in his emphasis on the *dimensions* that define them.

experimenter asks subjects to assign a number to the degree of similarity between two stimuli. We saw an example of this in the study of Terrace and Stevens (1962), who asked subjects to assign a number to the apparent volume of a test tone, given that a comparison tone was stipulated to have a volume of “10.” If we take differences in numerical assignments to represent the degree of volumetric dissimilarity between tones, then this experiment generates proximity data. Terrace and Stevens treated their subjects’ numerical assignments as legitimately quantitative, and thus the basis for a metric model of volume. More skeptically, however, we might treat them as merely revealing qualitative ordinal information about relative degrees of similarity: volumes of “10” fall between those of “3” and those of “17,” but this need not imply they are the “same” distance from the former as from the latter.¹⁷ A significantly more complex example of numerical proximity data is that of Dravnieks (1985), who asked subjects to rank odor samples on a scale of 0 to 5 for 146 different smell-related adjectives. This procedure produced for each odor stimulus a vector of 146 numbers, and two stimuli could be understood as more or less similar depending on how closely their vectors match. For instance, orange and lemon stimuli would both be ranked very high on the adjective “citrusy” and very low on “fishy.”

Shepard (1962a; 1962b) proposes multidimensional scaling as a method for extracting a low-dimensional metrical model from proximity data such as this, whether qualitative or numerical. The model produced by Shepard’s procedure is metrical in the sense that distances within it are meaningful, and if the data to which MDS is applied is psychophysical, the model it generates may be interpreted as a quality space. The decision to interpret the output of MDS as a quality space is a substantive one, in no way forced by the procedure itself; MDS may also be employed merely as a means to visualize data, without assigning any ontological interpretation to that visualization. However, it is the substantive interpretation that is of methodological interest, and which provides a helpful contrast with operationism (it is also the one that motivates Shepard himself). Although my examples are drawn exclusively from psychophysics, it is worth mentioning that MDS as a technique for manipulating data is employed in many other subfields of psychology, often with a substantive interpretation and epistemic attitude analogous to that discussed here.

Why would we want a method for representing proximity data in a low-dimensional space? The answer may not be obvious if we consider only data drawn from studies where subjects have been instructed to compare specific qualities like weight or tonal volume, as these are typically represented in low-dimensional spaces already, e.g. the curve of constant volume judgments through the 2-dimensional space of stimulus energy and frequency in Figure 1. The value of a procedure for producing low-dimensional representations of proximity data becomes more clear when we consider a case like Dravnieks’ data. Unlike that of tone stimuli, the physical description of odor stimuli (i.e. molecular structure) does not suggest any simple framework for organizing the data. The data themselves comprise unwieldy vectors of 146 numbers each. In order to represent the overall pattern of similarity between odors, then, we might plot each one in a 146-dimensional space, with each dimension representing the numerical assignments for a single adjective. In the general case, where our data is simply a list of pairwise assessments of similarity between n stimuli, we may need to represent it in a space of $n - 1$ dimensions (because then any point may be arranged at any distance whatsoever from the remaining $n - 1$). Large dimensional spaces such as this one, or that

¹⁷The method of “magnitude estimation,” or the direct assignments of numbers to stimuli by subjects, was developed by Stevens (1956) and has spawned its own literature on how best to interpret the data, in particular the question at issue here: whether it is qualitative or quantitative, and more specifically whether it should be represented by an ordinal, interval, or ratio scale (e.g. Krantz, 1972; Shepard, 1981). I don’t intend to take a stance in this debate here; the point is merely that one need not interpret such data as quantitative in order to subject it to MDS.

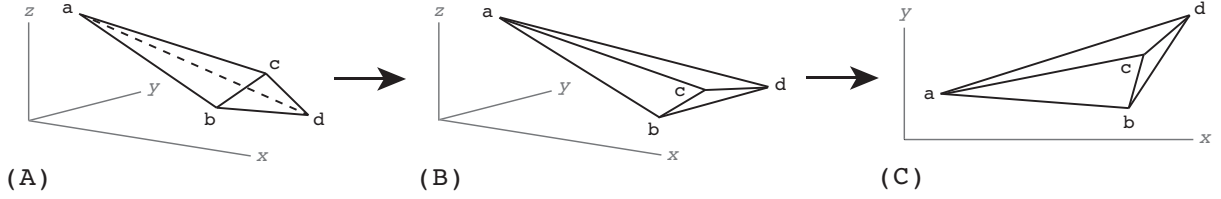


Figure 2: A toy example illustrating Shepard’s basic MDS algorithm. In **A**, $n = 4$ data points are arranged in $n - 1 = 3$ dimensions as an irregular tetrahedron, representing proximity distances ranked $ad > ac > ab > bd > cd > bc$; in **B**, the cluster of points has been stretched and flattened, preserving rank relative distance; in **C**, the axes have been rotated and points projected into two dimensions, again preserving relative distance.

of Dravnieks, are unwieldy, difficult to visualize or analyze for patterns, and of doubtful significance as representations of the psychological factors that ground subject similarity judgments. In contrast, low dimensional representations are both intuitively more parsimonious, and, under the right circumstances, mathematically unique (Shepard, 1962a, 131).

Shepard’s procedure aims to preserve the relative degrees of similarity, or distance, between elements while reducing the number of dimensions, until a representation of minimum dimensionality is found. If there are n data points, it first positions these within an $(n - 1)$ -dimensional space such that the proximities between them are all honored; it then iteratively reduces the dimensionality of the space in a manner that ensures the relative distances between points are preserved up to monotonicity. In other words, for any three data points a , b , and c , if a is closer to b than to c in the original space, then a is still closer to b than to c in the lower-dimensional space, even if the absolute distances between a , b , and c have all changed. Intuitively, the procedure stretches larger distances and compresses shorter ones, while preserving rank relative distance. This stretching and compressing essentially “flattens” the cluster of points overall. The coordinate system may then be rotated around the cluster until the largest portion of distances are projected onto the fewest number of axes. Other dimensions of the set are now projected to a single point on the remaining axes, the origin, rendering their contribution to the representation irrelevant. The “principal axes” may then be kept and the remainder discarded, leaving a new representation of the data in a space of lower dimensionality, and the procedure is applied again (Figure 2 depicts a toy example).

The iterative procedure Shepard proposes also allows for small departures from monotonicity, i.e. if the distance ab is greater than bc , but the two distances are very similar compared to other distances in the data set, the new representation may set $bc > ab$. Each representation in the iterated procedure is derived from the previous one by the joint application of two constraints: the first to maintain monotonicity with respect to distances in the original data; the second to “flatten” the overall arrangement of points. These constraints may be weighted differently, and if the second is weighted enough, small violations of monotonicity may arise. Since each iteration is constrained by the proximities in the original data set, however, a violation of monotonicity at one stage in the process may be corrected at a future stage. Once the iterative “flattening” and rotating steps no longer allow dimensions to be reduced, or monotonicity with respect to the original data to be more closely approached, the most faithful representation of minimal dimensionality has been found (Shepard, 1962a, 128–32).

Even if small violations in monotonicity, and thus faithfulness to the original data, are present in the final model, this might be judged an acceptable price to pay for achieving a dramatic simplification in its representation. If one suspected there to be noise in the data, or that its precision was otherwise compromised, one might even judge the representation in fewer dimensions a more accurate depiction of the underlying causal factors, an inference familiar from statistics:

Just as a regression line is usually a better predictor of further observations than the individual points upon which it is based, a spatial representation [of low dimensionality] can be both more reliable and more valid than the fallible data from which *it* was derived. (Shepard, 1966, 308)

Since Shepard’s procedure allows a quantitative assessment of the contribution of each axis to the preservation of distances, the exact nature of the tradeoff between faithfulness to the data and minimality of dimensions is explicit during analysis.

The representation of qualitative data in a metrical space may be perceived as a kind of sleight-of-hand, “tantamount to getting something for nothing” (1962b, 238). Shepard even characterizes his technique as “an automatic method for essentially transforming [a] given ‘ordinal scale’ of similarities into a ‘ratio scale’ of distances” (244). But a ratio scale contains more information than an ordinal scale—how could the latter be transformed into the former without some new data, or hidden, illegitimate assumptions? The key insight here is that metrical structure emerges from the compression of the ordinal information; it is not that information is added, but rather repurposed. In Shepard’s words:

[I]t is not that something is created out of nothing but, rather, that information distributed among many numbers in a dilute and inaccessible form is concentrated into a much smaller set of numbers where it can more readily be grasped and utilized. (239)

A low-dimensional representation of a high-dimensional data set requires many fewer parameters; consequently, each parameter bears more information. This is why “the degree of metric determinacy of the solution increases with the number of points” (239): as more parameters are compressed into fewer, the distances encoded in the fewer parameters become more meaningful. In the limit, with a sufficiently large number of data points, the low-dimensional representation approaches uniqueness up to a similarity transformation, i.e. multiplication by a constant. Since the characteristic feature of a ratio scale is that it is invariant across similarity transformations, this low-dimensional representation satisfies the formal properties of a metric representation, and distances within it are meaningful from a measurement theoretic standpoint (Shepard, 1966).¹⁸ These considerations underscore the value in finding the representation with the minimum number of dimensions: the lowest-dimensional representation extracts the maximum metric information from the data.

An example will illustrate how Shepard’s method may be used to determine which perceptual attributes are fundamental, in a manner that avoids the operationist worries of Section 2. Shepard (1962b) validates his method on both artificially generated data sets and actual psychophysical data, demonstrating it can recover antecedent structure or improve on previous analyses. One empirical data set he

¹⁸For the classic characterization of scales in terms of their invariance across transformations, see Krantz et al., 1971, Chapter 1; for the argument that the meaningfulness of aspects of a measurement scale should be equated with invariance across scale transformations, see Luce et al., 1990, Chapter 22. Shepard (1966) discusses the possibility of extracting metrical information from ordinal data in detail, providing both mathematical arguments for why it holds in the limit, and intuitive, empirical considerations for why it may obtain for real data.

considers was originally collected and analyzed by Ekman (1954), using an earlier method for reducing the dimensionality of data. At issue in these two analyses is the number and nature of the fundamental attributes of color experience, a worthy test given that this debate was one of those characterized by Boring (1942) as confused and ill-defined. Ekman showed 31 subjects two color stimuli at a time, windows covered in filters to produce single wavelength light, and asked them to report “the degree of ‘qualitative similarity’ on a scale with five steps, ranging from 0 (‘no similarity at all’) to 4 (‘identity’)” (468). Ekman sought the fewest number of dimensions characterizing the data, ultimately settling on five.¹⁹ These five underlying factors he identified with the “primary colors,” arguing that coordinates along each axis “indicate the relative contributions of the primary sensation factors to the complex sensation” (471). Since he had instructed subjects to merely attend to “qualitative similarity,” he interpreted this result as applying to hue only (and not brightness or saturation), concluding:

Our analysis, then, would indicate that five primary sensations, in different combinations, are responsible for all the more or less complex sensations of hue which are evoked by light stimulation of different wavelengths. (473)

Ekman’s result that five dimensions are required to reproduce hue similarity judgments is surprising given that we typically represent hues in two dimensions, arrayed around a circle. Shepard (1962b) applied his method to Ekman’s data. After two iterations he rotated to find the principal axes. “At that point the fractions of variance accounted for by the successive axes were as follows: .589, .250, .045, .041, .021, .019, .010, .008, . . . ” (236). Shepard then collapses all but two dimensions and continues the procedure, reasoning thus:

Since there was a marked drop after the second axis and since the first two axes together accounted for 84 per cent of the variance, iteration was continued in the two-dimensional space. The departure from monotonicity reached a rather low minimum of .00041 following iteration seven; thus the coordinates for that iteration were taken as the final solution. . . . [T]he degree of approach to a monotonic function shows that two dimensions are probably sufficient to represent the underlying structure of Ekman’s data. (236–7)

Two underlying dimensions or factors account almost completely for subject behavior on the perceptual task, and thus may be considered the fundamental attributes of hue experience, from which judgments of particular hue are derived. In fact, the arrangement of hues in Shepard’s final representation forms a circle, corresponding very closely to circular representations of hue derived through other methods.²⁰

¹⁹There is an abuse of terminology here in order to simplify the comparison with Shepard. Ekman’s method, Factor Analysis, looks for small numbers of unobserved variables that might explain correlations amongst large numbers of observed variables. Conceptually, this is very like the process of seeking a low-dimensional representation of distances (as Ekman himself indicates with the title of his paper), but the details of the method are radically different and suppressed here in the interests of space.

²⁰For another example of this kind of reasoning, consider the ongoing debate on the primitive attributes (or dimensions) of odor experience. Attempts to derive low-dimensional representations from odor proximity data date at least to Henning’s odor prism, proposed in 1916. Kaeppler and Mueller (2013) review a number of more recent attempts, including ones employing techniques such as those discussed here. Representative of one position in the debate is Koutrakov et al. (2011), who performed a principal component analysis of Dravnieks’ data, showing that 56% of variance in the data may be accommodated by a curved 2-dimensional surface embedded in three space, and an analogous curved surface in 10-dimensional space can account for upwards of 90%, motivating the view that odor may have as few as 10 fundamental attributes. In contrast, Young, Keller, and Rosenthal (2014) combine behavioral and physiological considerations to argue that a space of very high dimensionality is required to represent possible odor experience.

Notice the differences between this debate and that lampooned by Boring. Neither Shepard nor Ekman refer to their subjective impressions of simplicity or complexity, nor does Shepard’s critique appeal to intuition or vague concepts. Rather, the debate concerns the fitting of models to data, and the success of each approach can be assessed in quantitative terms. Notice also the contrast between this method and that of the operationists. True, Ekman was concerned with hue in his experiment, but he gave his subjects the very general instruction to evaluate “qualitative similarity,” not, as Stevens might mandate, to “attend to hue.” Nor was any training necessary for this simple task. Furthermore, Shepard’s circular representation was derived by an algorithmic procedure, and thus in no way influenced by his preconceptions about color and its metrical structure. As Shepard emphasizes, no assumption about the relationship between the original data and its final representation need be made other than that it preserve monotonicity. “The actual form of this function is left as one of the unknowns to be determined by the analysis itself” (219). This is the epistemic humility of multidimensional scaling: it does not impose the concepts of the experimenter on the subject, but extracts the underlying attributes of the subject’s experience through an impartial process.

In fact, we might worry that MDS is, if anything, too humble, this process too impartial. For instance, the procedure described above delivers a low-dimensional representation of data, but it does not offer any principled reason to prefer one orientation of the axes within this representation over any other. In the case of hue, Shepard’s analysis tells us there are two fundamental attributes, but it does not tell us what they are. Additional interpretation and justification would be required to identify these two attributes with, say, those of Hering: an axis of red–green opponency and one of blue–yellow opponency. As it stands, Shepard’s analysis is merely *consistent* with Hering’s. However, many methods for extracting low-dimensional representations from data are stronger than Shepard’s and *do* deliver a unique orientation of the axes, and thus specification of the fundamental attributes (for instance, Ekman’s own).²¹ Even when a method suggests specific candidate attributes, however, there is often work that must be done to make sense of its suggestion, i.e. to *interpret* the spatial structure in the model in terms that have perceptual or behavioral significance. So, for both MDS and operationism, the researcher’s antecedent theoretical commitments and intent enter into reasoning about the experiment, yet they do so at very different stages. For operationism, antecedent commitments and intent play a constitutive role *before* data is collected, when the subject is “tuned” to the attribute of interest through instruction and training. From the MDS perspective, prior commitments need only enter the process *after* both data collection and analysis, when the output of that analysis is interpreted.

Another point of contrast with operationism is in the role of physiological considerations. For Boring and Stevens, the physiology of sensation is the ultimate topic of perceptual psychology; consequently, any attempt to categorize attributes as more or less fundamental must rest ultimately on the nature of their physiological correlates. In contrast, the studies of Ekman and Shepard derive the fundamentality of attributes strictly from the behavioral data. Ekman discusses a possible physiological basis for his five primaries, and Shepard’s result is consistent with 2-dimensional, opponent-process theories of color physiology; nevertheless, these considerations do not drive their analyses. From the perspective of MDS, the physiology of sensation does not stipulate fundamentality; rather, the fundamentality of attributes is

²¹The INDSCAL algorithm of Carroll and Chang (1970) is the closest to Shepard’s that can deliver unique axes, and thus suggest specific fundamental attributes. Carroll and Chang add to Shepard’s method the assumption that subjects all employ the same criteria for similarity judgments, but weight them differently, allowing comparisons of data across subjects to privilege some orientations of the axes in a low-dimensional representation over others. This is the algorithm employed in the MDS studies on timbre discussed at the end of this section.

in the first instance a fact about experience, as accessed through behavior. The fundamentality of experiential attributes may, of course, provide clues to physiology, but success or failure at finding physiological correlates in no way undermines the legitimacy of the analysis.

Finally: what of tonal volume? The proceeding discussion should make clear that pitch, loudness, volume, and density would not be assigned equal importance in an MDS analysis. As Boring himself admitted, any two are enough to determine the remainder, and thus tone experience is fundamentally 2-dimensional. Or, rather, given the humility of MDS, the dimensionality of tone experience is yet to be determined, but of the four attributes identified by Stevens, only two of these should appear in any low-dimensional model of tonal attributes—most likely pitch and one of the other three, given the simple relationship amongst them identified by Stevens, Guirao, and Slawson (1965). Furthermore, the perceptual salience of loudness, and the need for extensive training in order to consistently judge volume, indicates that volume is not likely to emerge as a fundamental attribute in any experiment where subjects are merely instructed to report the similarity between tones.

But this is conjecture; what attributes emerge as fundamental if we apply the methods of MDS to proximity data on sounds? I'll close with two pertinent results. First, pitch itself turns out to be quite complex. While pitch experience might at first appear to be one dimensional, linearly ordered from low to high, this structure emerges only from experiments where simple stimuli are presented in random order. If, however, one plays a melody, or even simply a major scale, before presenting the subject with tones for comparison, new similarity relations emerge: tones separated by an octave are judged to be very similar, those separated by a fifth somewhat less so, but still more so than pitches that are linearly “nearby.” Shepard (1982) analyzes data gathered in a musical context such as this in order to motivate a five-dimensional model of musical pitch perception, the Euclidean product of the rectilinear scale from low to high, the two-dimensional circle of chroma, or pitch as identified independent of octave, and the two-dimensional circle of fifths (321). Second, MDS studies have more commonly been employed to study the perception of complex tones, stimuli characterized not only by frequency and energy, but also by the many overtones, or concurrent waves at different frequencies, that compose them. Typically, these studies control for pitch and loudness in order to study the perception of timbre, that further aspect of sounds by which we distinguish them. For instance, middle C's played on a piano, a violin, and a clarinet sound different, even when all are played at the same energy level—what perceptual attributes explain this judgment of difference? MDS provides a principled way to approach this question. An influential early study is Grey (1977), and McAdams et al. (1995) provide a survey and update. The results of such research remain ambiguous, with a single point of clarity: the one quality to definitively emerge as an attribute of sound experience in addition to loudness and pitch is *brightness*.

6 Conclusion: Attributes and Reality

Interest in tonal volume was once high, but now has waned. I have argued that the initial surge of interest in volume was provoked in part by an operationist methodology that constitutively involved the instruction and training of subjects. Operationism offered no criterion for evaluating the fundamentality of perceptual attributes other than conjectured physiological significance. Although several factors contributed to the decline of interest in tonal volume, I've focused on one in particular, the rise of multidimensional scaling and a host of similar methods exhibiting epistemic humility. These methods

algorithmically derive low-dimensional representations from high-dimensional data; the axes of these representations are then candidate fundamental attributes. Since the attributes are derived *ex post facto*, MDS does not require substantive instruction or training of subjects; since their fundamentality is latent in the perceptual data, it does not rest on physiological conjecture; and since the analysis is algorithmic, the expectations of the experimenter do not infect it.

A sweeping story such as this does not come without its caveats. The primary reason for the rise of MDS and related methods is not necessarily their epistemic features. A significant factor is the emergence of computers as a tool for data analysis. Shepard’s method could not have been seriously proposed before the early 1960’s as it is not feasible without computational assistance, and the increasing availability of computing power has driven the development of more elaborate methods in its stead. Furthermore, in emerging as possible techniques, MDS and its cousins have not driven training-heavy methods, such as those of Stevens, to extinction. The methods of operationism, behaviorism, and even reflective introspection all co-exist in the broad and heterogeneous practice of contemporary psychology. Nevertheless, I think the power to derive attributes rather than presuppose them has provoked a general skepticism toward over-training of subjects, and a broader sympathy for epistemic humility. While I have not provided comprehensive proof of a discipline-wide trend, I hope to have illustrated how it may manifest itself within the microcosm of auditory psychophysics.

Finally, even if I am right, and contemporary psychologists have increasingly embraced a norm of epistemic humility, humility is far from the only scientific virtue. In the limit, scientific norms come into conflict and must be traded off against each other for theory to progress. To mention just one example, the humility and local parsimony of MDS may stand in tension with norms of overall parsimony and theoretical coherence. Our overall theory of, say, audition should cohere across the domains of psychophysics, psychology, neuroscience, and physiology. The operationists took care to consider whether their approach would contribute to such coherence, providing a clear theoretical argument from the possibility of behavioral discrimination to the actuality of an underlying neural differentiation. It is not clear that such an argument exists for models derived by MDS. These models locally embody virtues such as parsimony and elegance that are expected to correlate with truth, and thereby support an inference to the neural reality of the fundamental dimensions they identify. But why expect parsimony as a feature of geometrical representations of behavioral data to coincide with parsimony as a feature of neurophysiological theories? Without a principled story about the relationship between low-dimensional models and physiology, MDS appears not much better off than operationism: the attributes it posits are surely “real” *in some sense*, but whether that reality is the appropriate one for a successful psychology of perception remains to be seen.

Acknowledgements

This paper has benefited from comments by J. E. Wolff, two anonymous reviewers, and participants at the Cambridge conference *The Making of Measurement*, July 23–4, 2015. It also owes an enormous debt to one of the anonymous reviewers for Isaac (2013), who first brought to my attention multidimensional scaling and its importance in recent psychophysics.

References

- Amundson, Ron (1986). “The Unknown Epistemology of E. C. Tolman”. In: *The British Psychological Society* 77.4, pp. 525–531.
- Barwich, Ann-Sophie and Hasok Chang (2015). “Sensory Measurements: Coordination and Standardization”. In: *Biological Theory* 10.3, pp. 200–211.
- Borg, Ingwer, Patrick J. F. Groenen, and Patrick Mair (2013). *Applied Multidimensional Scaling*. Heidelberg: Springer.
- Boring, Edwin G. (1921). “The Stimulus-Error”. In: *The American Journal of Psychology* XXXII.4, pp. 449–471.
- (1933). *The Physical Dimensions of Consciousness*. The Century Co.
- (1935). “The Relation of the Attributes of Sensation to the Dimensions of the Stimulus”. In: *Philosophy of Science* 2.2, pp. 236–245.
- (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton-Century-Crofts, Inc.
- Carroll, J. Douglas and Phipps Arabie (1998). “Multidimensional Scaling”. In: *Measurement, Judgment, and Decision Making*. Ed. by Michael H. Birnbaum. San Diego, CA: Academic Press, pp. 179–250.
- Carroll, J. Douglas and Jih-Jie Chang (1970). “Analysis of Individual Differences in Multidimensional Scaling via an N-way Generalization of “Eckart-Young” Decomposition”. In: *Psychometrika* 35.3, pp. 283–319.
- Carterette, Edward C. and Morton P. Friedman (1978). *Handbook of Perception vol. IV: Hearing*. New York, NY: Academic Press.
- Clark, Austen (1993). *Sensory Qualities*. Oxford: Clarendon Press.
- Coffin, Thomas E. (1941). *Some Conditions of Suggestion and Suggestibility: A Study of Certain Attitudinal and Situational Factors Influencing the Process of Suggestion*. Ed. by John F. Dashiell. Psychological Monographs 241 (vol. 53, no. 4). Evanston, IL: American Psychological Association, Inc.
- Ding, Cody S. (2013). “Multidimensional Scaling”. In: *The Oxford Handbook of Quantitative Methods in Psychology*. Ed. by Todd D. Little. Vol. 2: Statistical Analysis. Oxford: Oxford UP, pp. 235–256.
- Dravnieks, Andrew (1985). *Atlas of Odor Character Profiles*. Philadelphia, PA: ASTM.
- Ekman, Gosta (1954). “Dimensions of Color Vision”. In: *The Journal of Psychology* 38.2, pp. 467–474.
- Fechner, Gustav Theodor (1966 [1860]). *Elements of Psychophysics*. Ed. by Davis H. Howes and Edwin G. Boring. Trans. by Helmut E. Adler. New York, NY: Holt, Rinehart and Winston.
- Feest, Uljana (2005). “Operationism in Psychology: What the Debate is about, What the Debate Should be about”. In: *Journal of the History of the Behavioral Sciences* 41.2, pp. 131–149.
- Goldstein, E. Bruce (2007). *Sensation and Perception*. 7th. Belmont, CA: Thomson Learning Inc.
- Grassi, Massimo (2005). “Do we hear size or sound? Balls dropped on plates”. In: *Perception & Psychophysics* 67.2, pp. 274–284.
- Grey, John M. (1977). “Multidimensional Perceptual Scaling of Musical Timbres”. In: *Journal of the Acoustical Society of America* 61.5, pp. 1270–1277.
- Gulick, W. L. (1971). *Hearing: Physiology and Psychophysics*. New York, NY: Oxford University Press.
- Gulick, W. Lawrence, George A. Gescheider, and Robert D. Frisina (1989). *Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics*. New York, NY: Oxford UP.

- Gundlach, Ralph (1929). “Tonal Attributes and Frequency Theories of Hearing”. In: *Journal of Experimental Psychology* XII.3, pp. 187–196.
- Gundlach, Ralph and Madison Bentley (1930). “The Dependence of Tonal Attributes upon Phase”. In: *The American Journal of Psychology* 42.4, pp. 519–543.
- Halverson, H. M. (1922). “Diotic Tonal Volumes as a Function of Difference of Phase”. In: *The American Journal of Psychology* 33.4, pp. 526–534.
- (1924). “Tonal Volume as a Function of Intensity”. In: *The American Journal of Psychology* 35.3, pp. 360–367.
- Hardcastle, Gary L. (1995). “S. S. Stevens and the Origins of Operationism”. In: *Philosophy of Science* 62, pp. 404–424.
- Helmholtz, Hermann von (1962 [1910]). *Helmholtz’s Treatise on Physiological Optics*. Ed. by James P. C. Southall. 3rd ed. Vol. 2. New York, NY: Dover Publications, Inc.
- Hering, Ewald (1878). *Zur Lehre vom Lichtsinne*. Wein: Carl Gerold’s Sohn.
- Hollander, Ari J. (1994). “An Exploration of Virtual Auditory Shape Perception”. MA thesis. University of Washington.
- Isaac, Alistair M. C. (2013). “Quantifying the Subjective: Psychophysics and the Geometry of Color”. In: *Philosophical Psychology* 26 (2), pp. 207–233.
- James, William (1890). *Principles of Psychology*. 2 volumes. Henry Holt & Co.
- Kaepler, Kathrin and Friedrich Mueller (2013). “Odor Classification: A Review of Factors Influencing Perception-Based Odor Arrangements”. In: *Chemical Senses* 38, pp. 189–209.
- Koulakov, Alexei A. et al. (2011). “In Search of the Structure of Human Olfactory Space”. In: *Frontiers in Systems Neuroscience* 5.65, pp. 1–8.
- Krantz, David H. (1972). “A Theory of Magnitude Estimation and Cross-Modality Matching”. In: *Journal of Mathematical Psychology* 9, pp. 168–199.
- Krantz, David H. et al. (1971). *Foundations of Measurement*. Vol. 1. San Diego, CA: Academic Press, Inc.
- Kuehni, Rolf G. (2014). “Unique Hues and Their Stimuli—State of the Art”. In: *COLOR Research and Application* 39.3, pp. 279–287.
- Lenoir, Timothy (2006). “Operationalizing Kant: Manifolds, Models, and Mathematics in Helmholtz’s Theories of Perception”. In: *The Kantian Legacy in Nineteenth-Century Science*. Ed. by Michael Friedman and Alfred Nordmann. Cambridge, MA: MIT Press, pp. 141–210.
- Licklider, J. C. R. (1951). “Basic Correlates of the Auditory Stimulus”. In: *Handbook of Experimental Psychology*. Ed. by S. S. Stevens. New York, NY: John Wiley & Sons, Inc., pp. 985–1039.
- Luce, R. Duncan et al. (1990). *Foundations of Measurement*. Vol. 3. San Diego, CA: Academic Press, Inc.
- Marks, Lawrence E. and Daniel Algom (1998). “Psychophysical Scaling”. In: *Measurement, Judgment, and Decision Making*. Ed. by Michael H. Birnbaum. San Diego, CA: Academic Press, pp. 81–178.
- McAdams, Stephen et al. (1995). “Perceptual Scaling of Synthesized Musical Timbres: Common Dimensions, Specificities, and Latent Subject Classes”. In: *Psychology Research* 58, pp. 177–192.
- Michell, Joel (1997). “Quantitative Science and the Definition of *Measurement* in Psychology”. In: *British Journal of Psychology* 88, pp. 355–383.
- Moore, Brian C. J. (2008). *An Introduction to the Psychology of Hearing*. 5th. Bingley, UK: Emerald Group Publishing Limited.

- Odenthal, D. W. (1963). "Perception and Neural Representation of Simultaneous Dichotic Pure Tone Stimuli". In: *Acta Physiologica et Pharmacologica Neerlandica* 12, pp. 453–496.
- Perrott, David R. and Thomas N. Buell (1982). "Judgments of Sound Volume: Effects of Signal Duration, Level, and Interaural Characteristics on the Perceived Extensivity of Broadband Noise". In: *Journal of the Acoustic Society of America* 72.5, pp. 1413–1417.
- Perrott, David R., Alan Musicant, and Bettina Schwethelm (1980). "The Expanding-Image Effect: The Concept of Tonal Volume Revisited". In: *The Journal of Auditory Research* 20.1, pp. 43–55.
- Potard, Guillaume and Ian Burnett (2003). "A Study of Sound Source Apparent Shape and Wideness". In: *Proceedings of the 2003 International Conference on Auditory Display*, pp. 25–28.
- Rahn, Carl (1913). *The Relation of Sensation to other Categories in Contemporary Psychology*. The Psychological Monographs. Princeton, NJ: Psychological Review Company.
- Rich, G. J. (1916). "A Preliminary Study of Tonal Volume". In: *Journal of Experimental Psychology* 1.1, pp. 13–22.
- Rudno-Rudziński, K. and J. Renowski (1990). "Investigations on Auditory Spaciousness of Large Acoustic Sources within a Room". In: *Archives of Acoustics* 15 (3–4), pp. 447–463.
- Shepard, Roger N. (1962a). "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. I." In: *Psychometrika* 27.2, pp. 125–140.
- (1962b). "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. II." In: *Psychometrika* 27.3, pp. 219–246.
- (1966). "Metric Structures in Ordinal Data". In: *Journal of Mathematical Psychology* 3, pp. 287–315.
- (1981). "Psychological Relations and Psychophysical Scales: On the Status of "Direct" Psychophysical Measurement". In: *Journal of Mathematical Psychology* 24, pp. 21–57.
- (1982). "Geometrical Approximations to the Structure of Musical Pitch". In: *Psychological Review* 89.4, pp. 305–333.
- Stevens, S. S. (1934). "The Attributes of Tones". In: *Proceedings of the National Academy of Sciences of the United States of America* 20.7, pp. 457–459.
- (1935a). "The Operational Basis of Psychology". In: *The American Journal of Psychology* 47.2, pp. 323–330.
- (1935b). "The Operational Definition of Psychological Concepts". In: *Psychological Review* 42.6, pp. 517–527.
- (1956). "The Direction Estimation of Sensory Magnitudes—Loudness". In: *The American Journal of Psychology* 69.1, pp. 665–689.
- (1975). *Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects*. New York, NY: John Wiley & Sons.
- Stevens, S. S., Miguelina Guirao, and A. Wayne Slawson (1965). "Loudness, a Product of Volume Times Density". In: *Journal of Experimental Psychology* 69.5, pp. 503–510.
- Stumpf, Carl (1883). *Tonpsychologie*. Vol. 1. Leipzig: S. Hirzel.
- Takane, Yoshio, Sunho Jung, and Yuriko Oshima-Takane (2009). "Multidimensional Scaling". In: *The SAGE Handbook of Quantitative Methods in Psychology*. Ed. by Roger E. Millsap and Alberto Maydeu-Olivares. London: SAGE Publications Ltd., pp. 219–243.
- Terrace, H. S. and S. S. Stevens (1962). "The Quantification of Tonal Volume". In: *The American Journal of Psychology* 75.4, pp. 596–604.

- Thomas, Garth J. (1949). "Equal-Volume Judgments of Tones". In: *The American Journal of Psychology* 62.2, pp. 182–201.
- Titchener, Edward Bradford (1896). *An Outline of Psychology*. London: Macmillan & Co., Ltd.
- (1908). *Lectures on the Elementary Psychology of Feeling and Attention*. New York: The Macmillan Co.
- (1924). "The Term 'Attensity'". In: *The American Journal of Psychology* 35.1, p. 156.
- Tolman, Edward Chance (1926). "A Behavioristic Theory of Ideas". In: *Psychological Review* 33.5, pp. 352–369.
- (1958 [1936]). "Operational Behaviorism and Current Trends in Psychology". In: *Behavior and Psychological Man*. Berkeley: University of California Press, pp. 115–129.
- Turner, R. Steven (1994). *In the Mind's Eye*. Princeton, NJ: Princeton UP.
- von Békésy, Georg (1963). "Hearing Theories and Complex Sounds". In: *The Journal of the Acoustical Society of America* 35 (4), pp. 588–601.
- Watson, John B. (1913). "Psychology as the Behaviorist Views It". In: *Psychological Review* 20.2, pp. 158–177.
- Wright, Wayne (2016). "Psychologically Pure Colors". In: *Encyclopedia of Color Science and Technology*. Ed. by Ronnier Luo. Springer, ???
- Wundt, Wilhelm (1907). *Outlines of Psychology*. Trans. by Charles Hubbard Judd. 3rd ed. London: Wilhelm Engelmann.
- Young, Benjamin D., Andreas Keller, and David Rosenthal (2014). "Quality-Space Theory in Olfaction". In: *Frontiers in Psychology* 5.1, pp. 1–15.
- Zoll, Paul M. (1934). "The Relation of Tonal Volume, Intensity, and Pitch". In: *The American Journal of Psychology* 46.1, pp. 99–106.